A Comparison of Variational and Ensemble-Based Data Assimilation Systems for Reanalysis of Sparse Observations

Jeffrey S. Whitaker*

NOAA Earth System Research Laboratory/Physical Sciences Division, Boulder CO

Gilbert P. Compo

Climate Diagnostics Center/CIRES, University of Colorado, Boulder, CO and

NOAA Earth System Research Laboratory/Physical Sciences Division, Boulder CO

Jean-Noël Thépaut

ECMWF, Reading, UK

submitted to *Mon. Wea. Rev.*

November 6, 2008

*325 Broadway R/PSD1, Boulder, CO 80305-3328, Jeffrey.S.Whitaker@noaa.gov

Abstract

An observing system experiment, simulating a surface-only observing network representative of the 1930's, is carried out with three and four-dimensional variational data assimilation systems (3D-Var and 4D-Var) and an ensemble-based data assimilation system (EnsDA). It is found that 4D-Var and EnsDA systems produce analyses of comparable quality, and both are much more accurate than the analyses produced by the 3D-Var system. The EnsDA system also produces useful estimates of analysis error, which are not directly available from the variational systems.

1. Introduction

Compared to the observing networks used for modern operational numerical weather prediction (NWP), those used for historical reanalysis are often quite sparse. For example, prior to the advent of radiosondes in the 1940's, there were only a few hundred to a few thousand surface meteorological observing stations around the world. Currently, most operational centers assimilate hundreds of thousands to millions of individual observations, with remotely sensed data from space comprising the vast majority. Consequently, for historical reanalysis it is crucial to 'spread out' the information in the observations into unobserved regions and model state variables. In operational NWP, there is less need for this, since there are fewer model state variables that aren't strongly correlated with observations.

Observing networks for historical reanalysis are also quite inhomogeneous in time, varying from a few hundred synoptic surface observations in the early 20th century, to hundreds of thousands of surface, upper-air and remotely-sensed observations in the late 20th century (Bronniman et al. 2005). Therefore, analysis errors can change dramatically with time, and methods for assessing the impact of changing observation networks on analysis error are needed for climate studies that utilize long reanalysis datasets.

In this note, we examine how different data assimilation systems address these requirements. This is done by performing an observing system experiment, decimating the observations used for operational NWP in January and February 2005. The reduced set of observations are assimilated into a three-dimensional variational (3D-Var), a four-dimensional variational (4D-Var) and an ensemble data assimilation (EnsDA) system. The accuracy of the resulting analyses are assessed

by comparing to operational NWP analyses (which used all available observations). It is found that the 4D-Var and EnsDA systems produce analyses of comparable quality, and both are much more accurate than the analyses produced by the 3D-Var system. The EnsDA system also provides useful estimates of analysis error, which are not directly available from the 4D-Var system.

2. Experimental Design

a. The Observing System

Only surface pressure observations are assimilated, since reliable measurements are available for the entire 20th century, and they provide more information about the large-scale tropospheric circulation than observations of surface temperature or wind (Compo et al. 2006). The operational network of surface pressure observations in January 2005 is shown in Figure 1A. The operational surface network was degraded by including only observations from the Global Climate Observing System (GCOS) surface network over land, and only ship observations over water. This results in a 400% reduction in the total number of surface pressure observations, yielding a network of approximately 3800 observations per analysis time poleward of 20^{o}N (Figure 1B) that is similar to the network of digital observations that has been recovered for the 1930's (Whitaker et al. 2004, Figure 1).

b. The Data Assimilation Systems

The 3D-Var and 4D-Var assimilation experiments were run at ECMWF with the global data assimilation system and forecast model that were operational in May 2005 (cycle 29R1). The resolution of both the nonlinear forecast model and the tangent linear model was T159L60 (triangular truncation at total wavenumber 159 with 60 vertical levels). The 3D-Var systems uses FGAT (first-guess at analysis time) to make better use of asynoptic observations (Andersson et al. 1998). The 4D-Var system (Klinker et al. 2000) uses a 12-h window, and is similar to the system being used to generate the third-generation ECMWF global reanalysis (ERA-Interim). The EnsDA system is an implementation of the serial ensemble-square root filter, similar to that described in Whitaker et al. (2004) and Compo et al. (2006). In EnsDA, an ensemble of short-term forecasts from the previous analysis time is used to estimate the background-error covariances (e.g. Hamill 2006). The forecast model used to run the ensemble is the National Centers for Environmental Prediction (NCEP) global forecast model operational in June 2007, run at T62L28 resolution. The ensemble consists of 64 members.

c. Experiments

All three data assimilation systems were run from December 15, 2004 through February, 27 2005. Verification statistics were computed for January and February 2005 by comparing the resulting analyses to the operational NCEP 3D-Var analyses and ECMWF 4D-Var analyses, both of which used all available surface, upper-air and satellite observations. The background-error

variances for the ECMWF 3D-Var and 4D-Var systems were multiplied by factors of 7.29 and 4.0, respectively, to account for the reduced accuracy of the first-guess forecasts when only surface pressure observations are assimilated. These factors were computed using the fact that, if the assumptions inherent in the Kalman filter are satisfied, and the observation and background error covariances are optimal, then at the observation location the variance of the difference between observation and first-guess should equal the sum of the observation-error variance and background-error variance (e.g. Whitaker et al. 2008, equation 8). However, the 3D-Var and 4D-Var results shown here may not necessarily be optimal, since the background-error structure functions were not modified for the reduced network.

3. Results

Figure 2 shows time-series of Northern Hemisphere 500 hPa geopotential height and 850 hPa temperature root-mean-square error (relative to the NCEP operational analysis) for the three experiments. Both 4D-Var and the EnsDA produce analyses with less than half the error of 3D-Var. This result is not sensitive to the verifying analysis used, as the time mean error computed relative to both the NCEP operational analysis and an ECMWF 4D-Var analysis computed using all available surface, upper-air and satellite observations are very similar (Table 1). For 850 hPa temperature, and to a lesser extent 500 hPa geopotential height, the 4D-Var analysis has a small but significant advantage over EnsDA, although it is not clear to what extent this is due to the lower resolution of the model used in the EnsDA analysis, or to the differences in the data assimilation algorithms. Figure 3 shows an example set of 500 hPa geopotential height analyses for 12 UTC

February 20, 2005. Both 4D-Var and the EnsDA capture all of the synoptic scale features in the NCEP operational analysis, including the block over the North Atlantic. The 3D-Var analysis completely misses some large amplitude features, such as the cutoff low in the central North Pacific. The errors of the 4D-Var and EnsDA analyses are roughly comparable to 72 hour forecast errors in modern NWP systems (such as those used in the first-generation global reanalyses), consistent with the results of Compo et al. (2006). Consistent with previous results, these 3D-Var analyses are not significantly better than one could obtain with a purely statistical analysis system that does not use an NWP forecast model, but instead uses the climatological mean as a first-guess and climatological anomaly covariances for the background-error covariance (Whitaker et al. 2004; Compo et al. 2006; Kanamitsu and Hwang 2006; Bengtsson et al. 2004).

The superior performance of the 4D-Var and EnsDA systems can be traced to their ability to utilize flow-dependent background-error covariance estimates. In 4D-Var, the background-error covariance is evolved with the tangent-linear dynamics over the 12-hour assimilation window. In the EnsDA system, the background-error covariance is derived from a sample estimate using the 64-member forecast ensemble. Each ensemble member has its own analysis cycle, so that the background-error covariance information is continuously evolved as the ensemble is propagated through the analysis-forecast system. Also, by exploiting the time dimension in both the observations and the model background, 4D-Var and EnsDA can extract surface pressure tendency information from the observations and use this information to correct the model evolution. This capability does not exist with 3D-Var. Figure 4 shows the 500 hPa geopotential height analysis increments at 06 UTC February 1, 2005 produced by the EnsDA system for four hypothetical

observations, each 1 hPa greater than the ensemble mean background forecast, at four different locations along 45^oN, spaced 90^o apart. The increments vary considerably in scale, magnitude, and position relative to the observation, depending on the background synoptic flow characteristics. Figure 5 shows a comparison of the 4D-Var and EnsDA analysis increments for a single observation located at 45^oN and 160^oE. The 4D-Var increments are somewhat noisier than the EnsDA increments, possibly because the EnsDA system utilizes covariance localization (Hamill et al. 2001), while the 4D-Var system uses a wavelet-based formalism to compute and apply the background term in the cost function (Fisher 2005). Nevertheless, both systems produce increments that are maximum to the east of the observation location. suggesting that background-error covariances in both systems are reflecting the presence of a surface low and upper-level trough in the ensemble mean first-guess to the east of the observation location. In addition, the 4D-Var and EnsDA increments for this particular observation location are more similar to each other than the EnsDA increment is to EnsDA increments for other observation locations shown in Figure 4. This, together with the fact that the EnsDA and 4D-Var analysis accuracy are comparable (Figure 2), suggests that the flow-dependence in the background-error covariances captured in both systems is similar.

Analysis spread (standard deviation) provides a direct estimate of ensemble-mean analysis error in the EnsDA system. Such an estimate is not directly available from 4D-Var, although it can be approximated with an extra computation involving the inverse of the Hessian of the cost function (e.g Rabier and Courtier 1992, Appendix B). This is particularly important for reanalysis over very long periods, over which the observing network (and hence the analysis error) changes

substantially. Figure 6 illustrates this variation by showing both the ensemble mean and spread for 500 hPa height analyses for January 1, 1920 and 1950, obtained from the EnsDA system as part of the surface-pressure based NOAA-CIRES 20th Century Reanalysis Project. The solid contours in this figure show the ensemble mean, the shaded field is the analysis error estimate derived from the ensemble spread, and the dots show the location of the assimilated surface pressure observations. In 1920, the relatively sparse observing network is associated with a substantially larger expected analysis error, particularly over the Arctic and North Pacific.

4. Conclusions

Flow-dependent background error information provided by advanced four-dimensional assimilation schemes, such as 4D-Var and EnsDA, appear to be crucial in producing useful analyses when observations are very sparse, as is the case for meteorological observations over large regions of the globe prior to the widespread use of satellite data in the 1970's. Both 4D-Var and EnsDA perform comparably when given a network of sparse surface pressure observations similar to what is available for the 1930's. However, the EnsDA system does not require re-tuning of the background-error covariance model as the observing network changes. In the 3D-Var and 4D-Var experiments presented here, this re-tuning was done manually but could be automated based on evolving innovation statistics. Additionally, the EnsDA system directly provides an estimate of analysis error, while an expensive additional computation is required for 4D-Var. Estimates of analysis error may be important for the use of reanalysis datasets to test the statistical significance of analyzed changes in societally relevant quantities, such as storm frequency and intensity, over

periods in which the observing network changes substantially. More controlled experiments with 4D-Var and EnsDA (using not only the same set of observations, but the same forecast model) are needed to understand more fully the differences in the quality of the analyses produced by the two systems.

Acknowledgments

Fruitful discussions with P. Sardeshmukh, A. Simmons, N. Matsui and T. Hamill are gratefully acknowledged, as are the comments of two anonymous reviewers. This project would not have been possible without a U.S. Department of Energy Innovative and Novel Computational Impact on Theory and Experiment (INCITE) computing award at the National Energy Research Scientific Computer Center. Funding was provided by the NOAA Climate Program Office through the Climate Dynamics and Experimental Predictions (CDEP) program and through the Explaining Climate to Improve Predictions (ECIP) project.

References

Andersson, E., J. Haseler, P. Unden, P. Courtier, G. Kelly, D. Vasiljevic, C. Brankovica, C. Cardinali, C. Gaffard, A. Hollingsworth, C. Jakob, P. Janssen, E. Klinker, A. Lanzinger, M. Miller, F. Rabier, A. Simmons, B. Strauss, J.-N. Thepaut, and P. Viterbo, 1998: The ECMWF Implementation of Three-Dimensional Variational Assimilation (3D-Var). Part III: Experimental Results. *Quart. J. Roy. Meteor. Soc.*, **124**, 1831–1860.

Bengtsson, L., K. I. Hodges, and S. Hagemann, 2004: Sensitivity of the ERA40 reanalysis to the observing system: Determination of the global atmospheric circulation from reduced observations. *Tellus*, **56A**, 456–471.

Bronniman, S., G. Compo, P. Sardeshmukh, R. Jenne, and A. Steri, 2005: New approaches for extending the Twentieth Century climate record. *Eos, Trans. Amer. Geophys. Union*, **86**, 2–6.

Compo, G. P., J. S. Whitaker, and P. D. Sardeshmukh, 2006: Feasibility of a 100 year reanalysis using only surface pressure data. *Bull. Amer. Meteor. Soc.*, **87**, 175–190.

Fisher, M., 2005: Background-error covariance modelling. *Seminar on recent developments in data assimilation for atmosphere and ocean*, ECMWF, Reading, UK, 45–63.

Hamill, T. M., 2006: Ensemble-based data assimilation. *Predicability of Weather and Climate*, Cambridge Press, chapter 6, 124–156.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790.

Kanamitsu, M. and S. O. Hwang, 2006: The role of sea-surface temperatures in reanalysis. *Mon. Wea. Rev.*, **134**, 532–552.

Klinker, E., F. Rabier, G. Kelly, and J.-F. Mahfouf, 2000: The ECMWF Operational Implementation of Four-Dimensional Variatonal Assimilation. Part III: Experimental results and diagnostics with operational configuration. *Quart. J. Roy. Meteor. Soc.*, **126**.

Rabier, F. and P. Courtier, 1992: Four-Dimensional Assimilation In the Presence of Baroclinic Instability. *Quarterly Journal of the Royal Meteorological Society*, **118**, 649–672.

Whitaker, J. S., G. P. Compo, and T. M. Hamill, 2004: Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.*, **132**, 1190–1200.

Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble Data Assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **138**, 463–482.

Figure Captions

Figure 1: Operational (A) and thinned (B) surface pressure networks for 00 UTC 1 January 2005

Figure 2: Time series of Northern Hemisphere (poleward of 20^o N) 500 hPa geopotential height (top panel) and 850 hPa temperature (lower panel) root-mean-square analysis error (measured relative to NCEP operational analysis) for 3D-Var (black), 4D-Var (blue) and the EnsDA (red). The 4D-Var and 3D-Var analyses are from the ECMWF system at T159L60 resolution (triangular truncation at wavenumber 159 with 60 vertical levels), while the EnsDA analysis uses the NCEP global model at T62L28 resolution.

Figure 3: Example 500 hPa geopotential height analyses for 12 UTC February 20, 2005. Contour interval 50 m. The lower right panel shows the NCEP operational analyses, which used all available observations, and is used as a reference to estimate analysis error. The root-mean square analysis error in the Northern Hemisphere poleward of 20^oN is noted on each panel.

Figure 4: Analysis increments for geopotential height at 500 hPa for four hypothetical single observations located along 45^oN, spaced 90^o apart. The analysis increments are computed using the first-guess ensemble from the EnsDA system at 06 UTC 1 February 2005. Each observation is assumed to be 1 hPa higher than the ensemble mean first-guess surface pressure. The thin color-filled contours (interval 1 m, zero contour suppressed, red shading positive, blue shading negative) are the analysis increments, while the thicker contours (not color-filled) are the ensemble mean first-guess forecast (interval 50 m).

Figure 5: Analysis increments for geopotential height at 500 and 1000 hPa for a single hypothetical single observations located along 45^oN and 160^oE, for both the EnsDA and 4D-Var systems at 06 UTC 1 February 2005. The observation is assumed to be 1 hPa higher than the ensemble mean first-guess surface pressure. The thin color-filled contours (interval 1 m for 500 hPa, 2 m for 1000 hPa, zero contour suppressed, red shading positive, blue shading negative) are the analysis increments, while the thicker contours (not color-filled, zero line thickened) are the ensemble mean first-guess forecast (interval 50 m).

Figure 6: Ensemble mean analysis (solid contours) and analysis spread (color shading) for 500 hPa geopotential height analyses on 00 UTC January 1, 1920 and 1950. The contour interval is 50 m for the ensemble mean, with the 5600 m contour thickened. The color-scale for the analysis spread is shown on the right of each panel in meters. The red dots denote the position of all the surface pressure observations assimilated at each analysis time.

Table 1: 500 hPa geopotential height and 850 hPa temperature root-mean-square analysis errors measured relative to both the NCEP operational 3D-Var analysis and an ECMWF 4D-Var analysis using all available observations. The error relative to the ECMWF analysis is in parentheses. The errors are computed for the 3D-Var, 4D-Var and EnsDA systems over the Northern Hemisphere extratropics (poleward of 20^oN).

	EnsDA	4D-Var	3D-Var
500 hPa height (m)	33.47 *(32.75)*	31.24 *(29.87)*	111.67 *(111.10)*
850 hPa temperature (K)	2.60 *(2.58)*	2.27 *(2.02)*	5.67 *(5.62)*

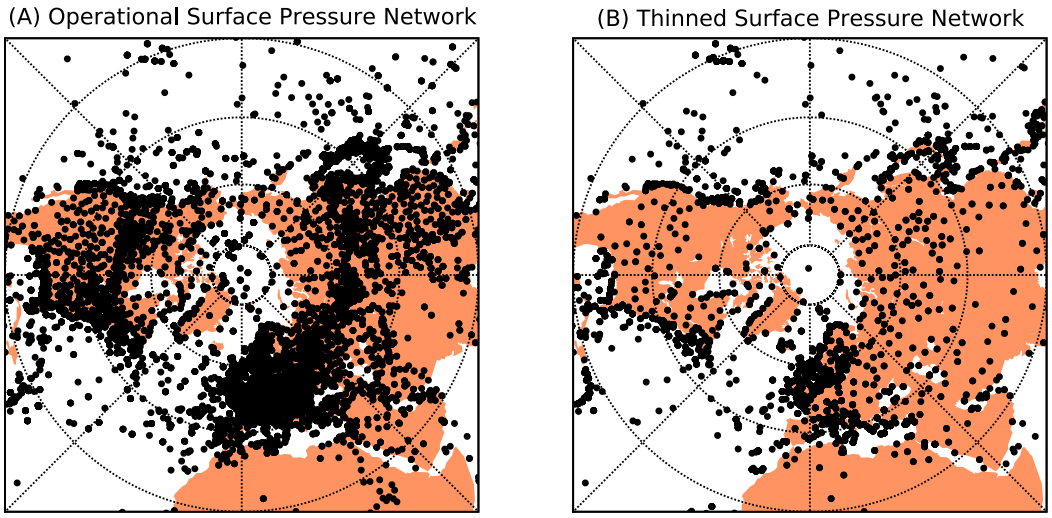

Figure 1: Operational (A) and thinned (B) surface pressure networks for 00 UTC 1 January 2005

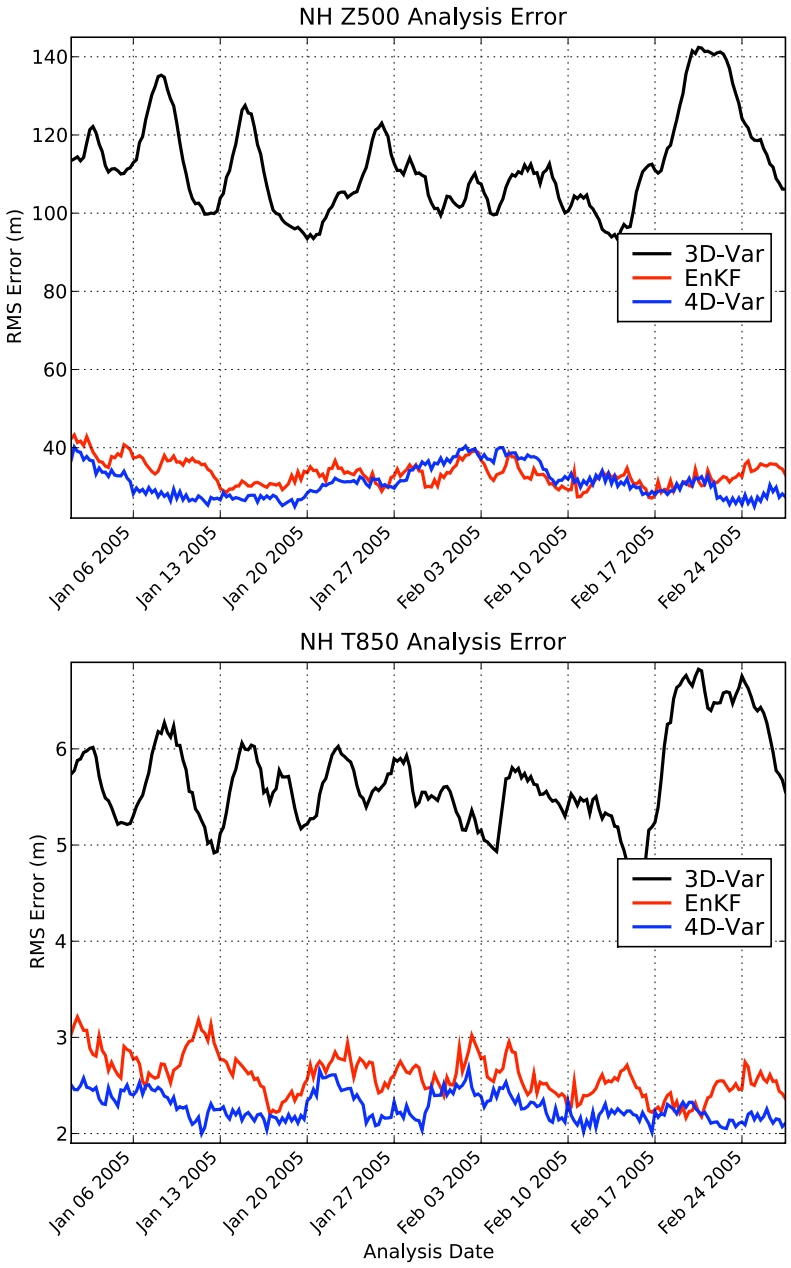

Figure 2: Time series of Northern Hemisphere (poleward of 20^o N) 500 hPa geopotential height (top panel) and 850 hPa temperature (lower panel) root-mean-square analysis error (measured relative to NCEP operational analysis) for 3D-Var (black), 4D-Var (blue) and the EnsDA (red). The 4D-Var and 3D-Var analyses are from the ECMWF system at T159L60 resolution (triangular truncation at wavenumber 159 with 60 vertical levels), while the EnsDA analysis uses the NCEP global model at T62L28 resolution.

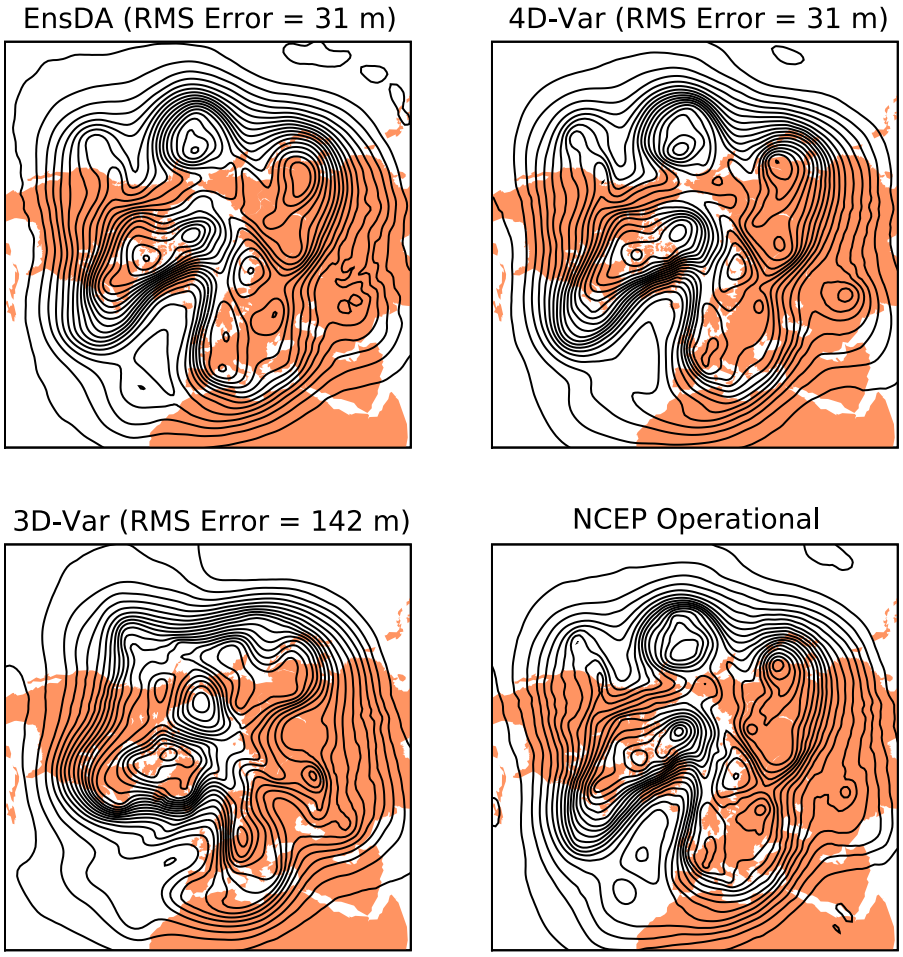

Figure 3: Example 500 hPa geopotential height analyses for 12 UTC February 20, 2005. Contour interval 50 m. The lower right panel shows the NCEP operational analyses, which used all available observations, and is used as a reference to estimate analysis error. The root-mean square analysis error in the Northern Hemisphere poleward of 20^oN is noted on each panel.

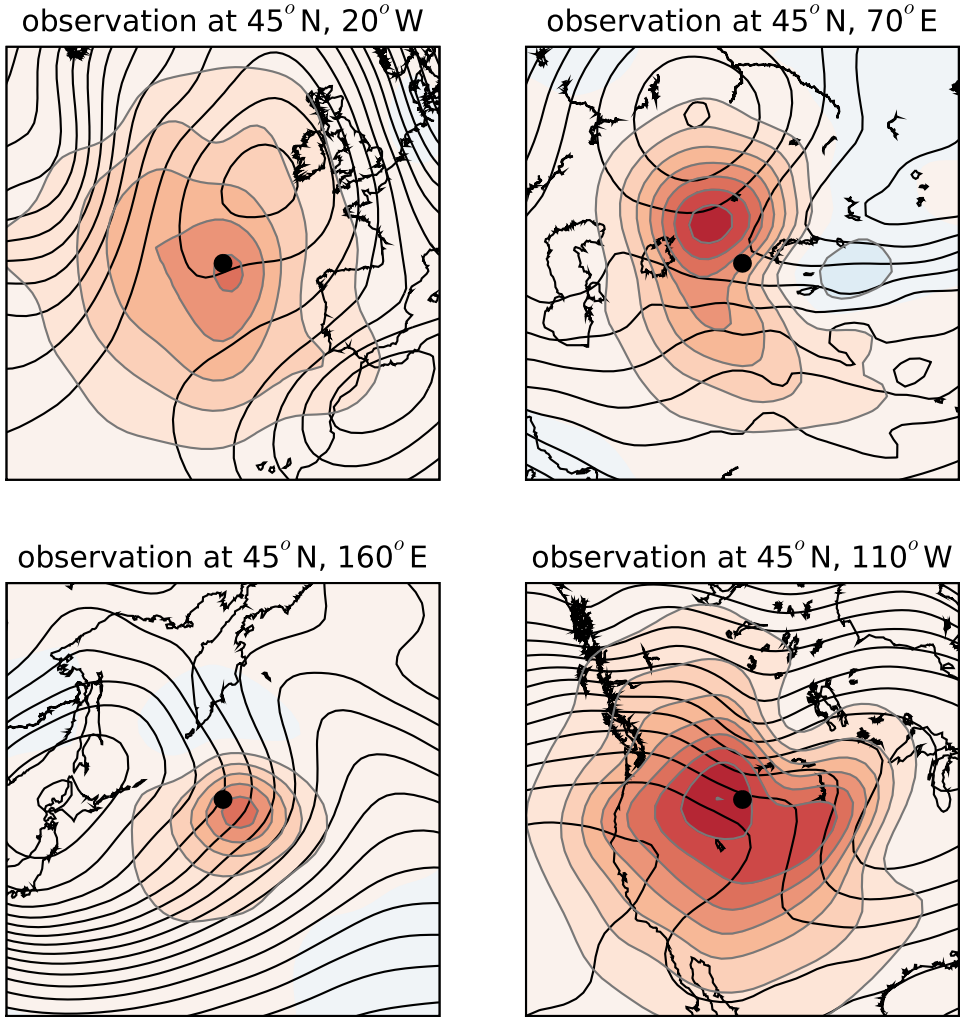

Figure 4: Analysis increments for geopotential height at 500 hPa for four hypothetical single observations located along 45^{o}N, spaced 90^o apart. The analysis increments are computed using the first-guess ensemble from the EnsDA system at 06 UTC 1 February 2005. Each observation is assumed to be 1 hPa higher than the ensemble mean first-guess surface pressure. The thin color-filled contours (interval 1 m, zero contour suppressed, red shading positive, blue shading negative) are the analysis increments, while the thicker contours (not color-filled) are the ensemble mean first-guess forecast (interval 50 m).

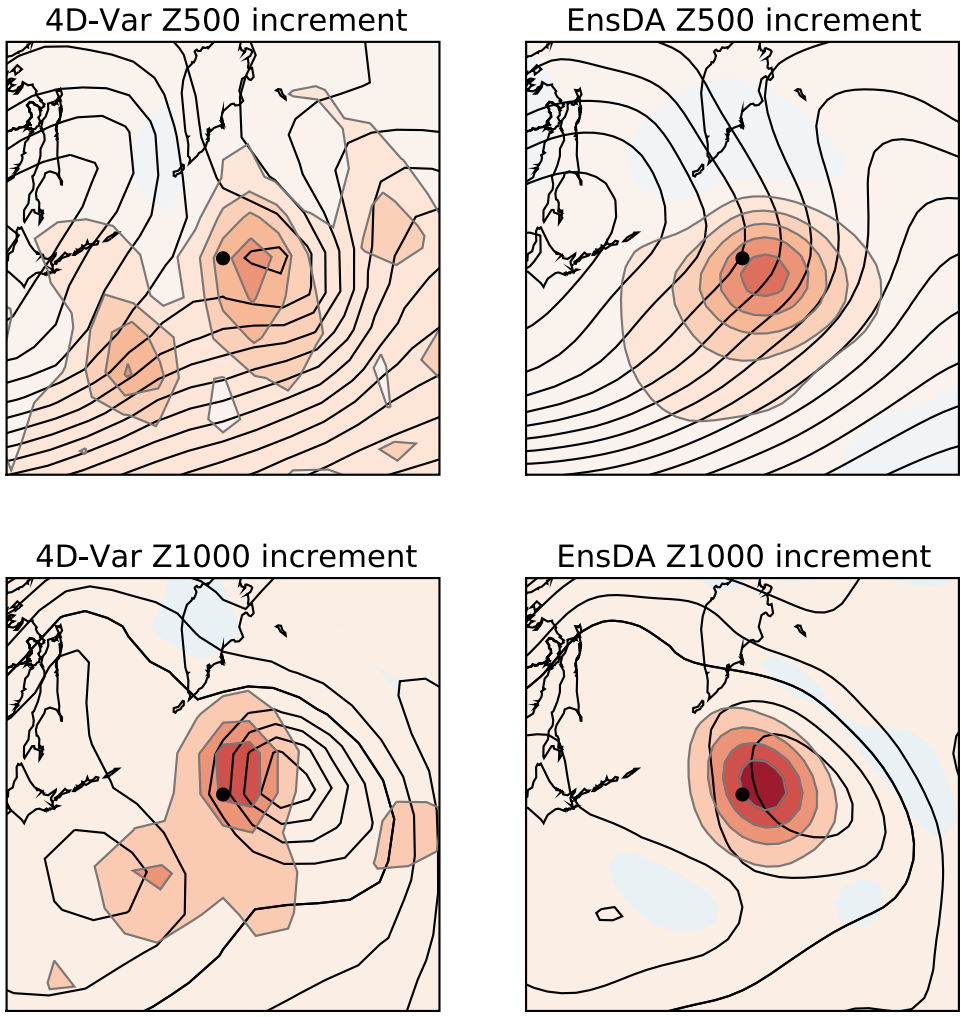

Figure 5: Analysis increments for geopotential height at 500 and 1000 hPa for a single hypothetical single observations located along 45^oN and 160^oE, for both the EnsDA and 4D-Var systems at 06 UTC 1 February 2005. The observation is assumed to be 1 hPa higher than the ensemble mean first-guess surface pressure. The thin color-filled contours (interval 1 m for 500 hPa, 2 m for 1000 hPa, zero contour suppressed, red shading positive, blue shading negative) are the analysis increments, while the thicker contours (not color-filled, zero line thickened) are the ensemble mean first-guess forecast (interval 50 m).

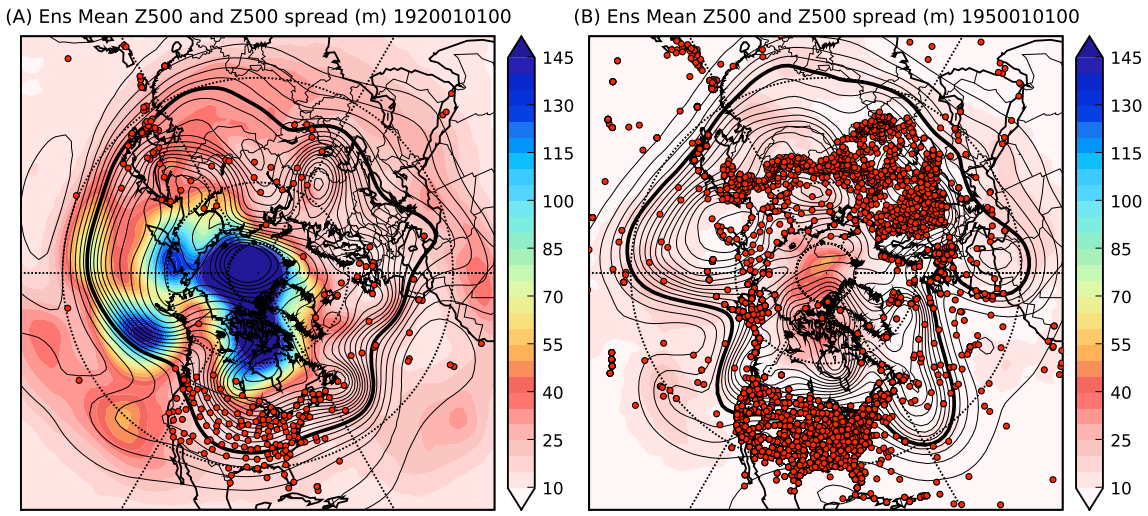

Figure 6: Ensemble mean analysis (solid contours) and analysis spread (color shading) for 500 hPa geopotential height analyses on 00 UTC January 1, 1920 and 1950. The contour interval is 50 m for the ensemble mean, with the 5600 m contour thickened. The color-scale for the analysis spread is shown on the right of each panel in meters. The red dots denote the position of all the surface pressure observations assimilated at each analysis time.